

4

SISTEMA DE GESTIÓN DEL ARCHIVO DE IMAGEN Y SONIDO DEL SENADO

José Ángel Alonso López

Jefe del Área de Desarrollo de la Dirección de Informática de la Secretaría General
Senado

PRECEDENTES

En noviembre de 1999, dos años después de la inauguración de la página web del Senado, se iniciaron las pruebas de emisión en directo mediante streaming de la señal de vídeo de las sesiones plenarias. Al comienzo de la VII Legislatura (marzo 2000) se normalizó el servicio, que incluía también el almacenamiento en la web a disposición del usuario de los vídeos de los Plenos más recientes.

Este servicio demostró desde el primer momento su utilidad y permitió obtener una importante experiencia sobre la gestión multimedia, así como ideas claras sobre los requisitos de un sistema de este tipo. En concreto, quedó claro lo siguiente:

- La web es el sitio idóneo para prestar este servicio.
- Además de transmisiones en directo era muy útil el almacenamiento de vídeos de sesiones pasadas, pero siempre que fuera posible acceder directamente al asunto o, incluso, a la intervención concreta requerida.
- Era necesario una profunda imbricación entre los vídeos y la información de la actividad parlamentaria disponible en la web del Senado.
- El volumen de información a tratar y almacenar requería disponer de unos potentes equipos dedicados exclusivamente a este fin.

Durante 2002 el Senado estudió la conveniencia de dotar a sus dependencias de un sistema de captación, tratamiento y grabación de sesiones. El objetivo era facilitar a los medios de comunicación una señal institucional, que también se transmitiría a través del satélite Hispasat a las plataformas de televisión digital. Para obtener el equipamiento necesario en los dos edificios del Palacio del Senado (cámaras robotizadas, equipos de realización, etc.) se llevó a cabo el correspondiente concurso. El material indicado se incorporó a un primer lote del pliego del concurso.

Por otro lado, se definieron las características de un sistema completo para la captura y almacenamiento en formato digital de material de audio y vídeo, así como para su posterior tratamiento y explotación.

Teniendo en cuenta la concurrencia de actos simultáneos, el sistema debía permitir la captura, catalogación e indexación de cuatro señales simultáneas de vídeo y audio. Asimismo, la capacidad inicial del sistema debía permitir almacenar el material audiovisual generado en una Legislatura. El sistema debía imbricarse profundamente con el **sistema de gestión de iniciativas parlamentarias de la Cámara (Gelabert)** y con la página web del Senado, permitiendo obtener los vídeos correspondientes a las sesiones completas, así como los correspondientes a la discusión de asuntos concretos e, incluso, intervenciones individuales de cualquier orador.

El sistema debía almacenar vídeo en formato digital de alta calidad para su entrega a medios de comunicación y reproducción en soporte DVD, CD-R, etc. También se requería el almacenamiento en formato de baja calidad para su emisión utilizando técnicas de streaming por Internet.

El equipamiento requerido integró el segundo lote del citado concurso que, tras su publicación en el verano de 2002, fue adjudicado en noviembre de ese mismo año. El lote, que se adjudicó a la empresa Promovisa, incluyó el software Tarsys para la gestión integral del sistema.

DESCRIPCIÓN DEL EQUIPAMIENTO ADQUIRIDO

El sistema recibe simultáneamente hasta cuatro fuentes de vídeo y audio de alta calidad (formato SDI). Por un lado se procesan estas señales en dos sistemas Pentium IV a 2,4 Ghz., con 512 Mb RAM, sistema operativo Windows 2000 y tarjeta de captura de audio y vídeo Osprey, generando vídeo en formato Real para streaming en Internet, con un baud rate 150 Kb/seg. Esta información es almacenada en otro servidor que es el que sirve por Internet los que se le solicitan (Compaq Proliant DL380, biprocesador Pentium III a 1,4 Ghz, con 768 Mb. RAM y 34,8 Gb de disco). Este sistema está dotado de sistema operativo LINUX Red Hat 7.2.

Por otro lado, las señales de vídeo en formato SDI son tratadas por dos servidores de captura de vídeo dotados con tarjetas Matrox Digiserver de 2 entradas y 4 salidas y tarjetas Optibase Moviemaker 100S Plus para codificación en MPEG-1. Estos sistemas son Compaq Proliant biprocesador pentium III a 2,4 Ghz, con 1 Gb RAM y 72,8 Gb de disco duro. Disponen de sistema operativo windows 2000. En ellos se generan los siguientes formatos:

- MPEG-1, formato de baja calidad de 1,2 Mbits/seg. Este material es el que se utiliza por el sistema de catalogación.
- MPEG-2, formato de alta calidad de 12 Mbits/seg. Este material es el que se pondrá a disposición de los usuarios que lo requieran (televisiones, etc.).

El almacenamiento de toda esta información requiere un subsistema de gran capacidad, que consta de los siguientes elementos:

- Sistema gestor de almacenamiento y transporte (AST). Se encarga de la gestión y la transferencia entre los servidores de contenidos la estación S.A.N. (Storage Area Network) y la librería de cintas, proporcionando un almacenamiento "buffer" a estos sistemas para asegurar una velocidad constante en las lecturas y las escrituras. Se trata de un Compaq Proliant DL380, biprocesador Pentium III a 1,4 Ghz, con 768 Mb. RAM y 328,4 Gb de disco. Este sistema está dotado de sistema operativo LINUX Red Hat 7.2.
- Estación S.A.N. MSA 1000, con capacidad de 1,5 Tbytes. Permite almacenar 2.500 horas de vídeo en formato "Real" para streaming por Internet, y 25 horas de vídeo en formato MPEG-2.
- Librería de cintas StorageTek.L700 con capacidad para 8 Tbytes en 180 cintas LTO de 100 gbytes cada una. Esto debe permitirnos almacenar más de 3000 horas de vídeo en formato MPEG-2.

Por otro lado, Tarsys reside en un servidor Compaq Proliant ML530, biprocesador Pentium III a 2,4 Ghz, con 1 Gb. RAM y 182,4 Gb de disco. Este sistema está dotado de sistema operativo LINUX Red Hat 7.2. La aplicación Tarsys controla todo el sistema y dispone de un gestor de contenidos, el gestor de base de datos ORACLE, la base de datos del sistema y el archivo de storyboards.

Por último, se dispone de un sistema que permite generar materiales en formato MPEG-1 para su almacenamiento en el robot. Se trata de un Compaq Proliant DL380, biprocesador Pentium III a 1,4 Ghz, con 786 MB. RAM y 328,4 Gb de disco duro. Este sistema está dotado también de sistema operativo LINUX Red Hat 7.2.

Con todos estos ordenadores se ha formado una red independiente Gigabit ethernet conectada mediante un Switch Cisco WS 3550 24 T.

INTEGRACIÓN DEL SISTEMA DE ARCHIVO DE IMAGEN Y SONIDO CON GELABERT

El sistema de gestión de la tramitación de expedientes parlamentarios realiza el seguimiento completo desde la entrada en el registro de la Cámara hasta la finalización de cada asunto. Algunos de los actos más importantes de esta tramitación se sustancian en sesiones plenarias o de comisión. Esto significa que la información sobre lo que se va a debatir en las sesiones, su orden del día, se encuentra previamente en Gelabert. Por otro lado, al final de las sesiones, el sistema de vídeo dispone de información de lo tratado que interesa tener en Gelabert.

Por otro lado, el sistema de vídeo dispone continuamente de información sobre quien está haciendo uso de la palabra, y sobre qué habla. Estos dos datos deberían transmitirse a Gelabert para subtítular las imágenes que se emitan en directo a través de la web del Senado. Todo lo anterior justificaba la conveniencia de realizar un trabajo de integración entre ambos sistemas. Para ello, se comenzó por definir el nivel de detalle de los metadatos que acompañarían a los materiales multimedia. Dentro de cada sesión se discute sobre distintos asuntos parlamentarios: comparecencias de ministros y altos cargos, interpelaciones, mociones, proyectos de ley, etc. Por tanto, es el asunto el segundo nivel de información que debe considerarse. Sin embargo, la tramitación de determinados asuntos, por ejemplo proyectos de ley, puede alargarse durante horas. La manipulación de vídeos de este tamaño es poco ágil y no permite acceder al usuario al ítem concreto que normalmente se desea: esos pocos minutos en los que se produce una declaración política de importancia o un intercambio de opiniones especialmente significativo entre responsables políticos. Era necesario definir un tercer nivel de información: la intervención.

Se definieron dos mecanismos para intercambio de información entre ambos sistemas: tablas y vistas de base de datos y ficheros XML transmitidos desde el sistema Tarsys a Gelabert.

Los dos sistemas comparten el mismo gestor de base de datos (ORACLE), mediante el cual se ha implementado un mecanismo de comunicación bidireccional. Se definen un conjunto de 24 vistas en Gelabert, que permiten a Tarsys obtener toda la información sobre los órdenes del día de las sesiones que se van a producir, así como múltiple información necesaria para su gestión: datos de salas, tipos de expedientes, senadores, otros intervinientes, fases de tramitación, etc.

Se crearon también tres tablas para almacenar información de las sesiones, asuntos e intervenciones producidas. Estos datos son insertados desde Tarsys en la base de datos de Gelabert, cuando finaliza cada sesión.

Sin embargo, este mecanismo no permitía cubrir adecuadamente la edición de información en la página web durante la sesión. Además, se consideró interesante disponer de información de la misma en formato XML, ya que podía ser publicada inmediatamente en Internet usando las correspondientes hojas de estilo. Para ello se implementó un segundo mecanismo consistente en la transmisión desde Tarsys a Gelabert de ficheros con información de toda la sesión, de lo que sucedía en cada momento y otra información específica de cada corte de vídeo. Esta transmisión se realiza mediante un programa CGI residente en una máquina servidora de ficheros de la red general del Senado (en la que se encuentra Gelabert).

FICHEROS QUE SE TRANSMITEN DURANTE LA EMISIÓN EN DIRECTO

Con ellos Tarsys informa a Gelabert de lo que está pasando en la sesión en cada momento (quien habla y sobre qué) y se detallan los asuntos anteriores ya terminados de la sesión que se transmite en directo, a cuyos vídeos se puede acceder mientras continúe la emisión.

directo_temp_XXXXXXX_yyy_zz.xml
diferido_temp_XXXXXXX_yyy_zz.xml

Interviniente y asunto del que habla.

Con un formato análogo al anterior almacena información sobre los asuntos terminados de la sesión que se está emitiendo en directo. Esto permite que el usuario que está viendo una emisión en directo pueda ver el corte de vídeo de una intervención que terminó hace horas de la sesión en curso.

asu_temp_XXXXXXX_yyy_zz_aaa.ram
asu_temp_XXXXXXX_yyy_zz_aaa.smi
int_temp_XXXXXXX_yyy_zz_aaa_bbb.ram
int_temp_XXXXXXX_yyy_zz_aaa_bbb.smi
real_XXXXXXX_yyy_zz_nnn.rm

Tantos como asuntos terminados haya en la sesión actual.

Tantos como intervenciones terminadas haya en la sesión actual.

Fichero de 30 minutos, tantos como sean necesarios para recoger los vídeos de la sesión.

El detalle del contenido de estos ficheros y de los componentes de sus nombres se incluye en el anexo I.

FICHEROS QUE SE TRANSMITEN TRAS LA FINALIZACIÓN DE LA SESIÓN

Contienen el detalle de la sesión, de los asuntos que se trataron y de todas las intervenciones realizadas, para su visualización en la web.

diferido_fin_XXXXXXX_yyy_zz.xml

Solo uno para toda la sesión, con información general de la misma, de sus asuntos e intervenciones una vez terminada la emisión en directo.

Uno para cada sesión.

ses_fin_XXXXXXX_yyy_zz.ram
ses_fin_XXXXXXX_yyy_zz.smi
asu_fin_XXXXXXX_yyy_zz_aaa.ram
asu_fin_XXXXXXX_yyy_zz_aaa.smi
int_fin_XXXXXXX_yyy_zz_aaa_bbb.ram
int_fin_XXXXXXX_yyy_zz_aaa_bbb.smi
real_XXXXXXX_yyy_zz_nnn.rm

Tantos como asuntos para su consulta en la web.

Tantos como intervenciones para consulta en web.

Fichero de 30 minutos, tantos como sean necesarios para recoger los vídeos de la sesión .

El detalle del contenido de estos ficheros y de los componentes de sus nombres se incluye también en el anexo I.

PROCESO “EN DIRECTO”

El sistema permite la ingesta de las señales de cuatro sesiones simultáneas. Durante el proceso, el operador incorpora los metadatos necesarios para poder acceder a la información poste-

riormente y se transmite la sesión en directo con información complementaria (el proceso en detalle se describe en el esquema del anexo II).

Las señales llegan en formato SDI (vídeo profesional) y se distribuyen a los servidores de codificación a formato streaming y a los servidores de captura.

Los servidores de captura (SENCPI52 y SENCPI54) disponen cada uno de dos tarjetas para la codificación a formatos MPEG-1 y MPEG-2. Los materiales en formatos MPEG-2 (12 Mbits/seg) son entregados al servidor de almacenamiento y transporte, que se encargará de que las 25 horas más recientes en este formato permanezcan en el sistema S.A.N., por si es necesario disponer de ellas rápidamente, pasando los materiales más antiguos a la librería de cintas StorageTek. El vídeo en formato MPEG-1 se entrega al sistema de almacenamiento SENCPI53 donde permanecen las 400 horas más recientes en este formato. Por otro lado, se entregan al cliente de captura que corresponda, donde el operador que ha iniciado el proceso trabaja en la pantalla de la figura 1.

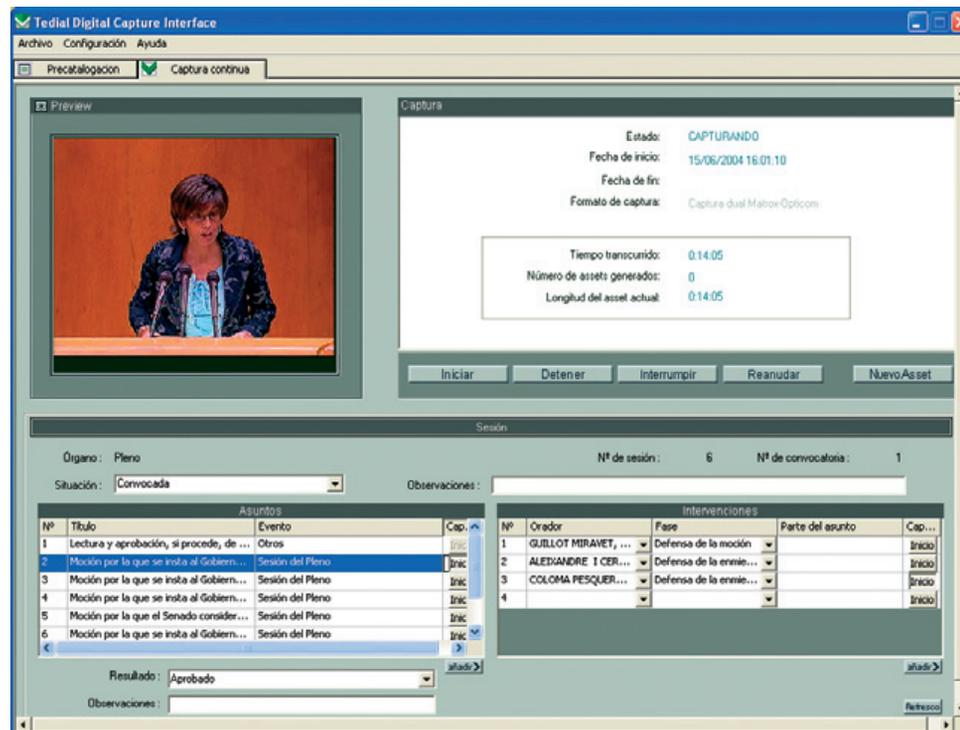


Figura 1

Además de la imagen de la sesión que aparece en la zona superior izquierda, el operador dispone de información sobre el estado del proceso de ingesta (zona superior derecha), y de la información de la sesión en la zona inferior. En la parte izquierda están los asuntos que se tratan en la sesión, que habrán sido cargados previamente desde Gelabert, aunque es posible incorporar más “sobre la marcha”. En la parte derecha, el operador irá incorporando la información sobre los oradores que toman la palabra, y el motivo de su intervención (defender una moción,

explicar la posición de un grupo parlamentario, explicar el voto negativo a un artículo de un proyecto de ley, etc.). Es posible indicar, en los asuntos complejos, la parte del asunto que se trata en cada momento. Por ejemplo, en enmiendas a un artículo de un proyecto de ley.

Estos metadatos van siendo recogidos por Tarsys, que los usará para dos fines:

- Asociarlos a los cortes de vídeo en el archivo de imagen y sonido.
- Incorporarlos a ficheros que se van a transmitir a Gelabert, tal como se ha indicado anteriormente.

Por otro lado, las señales en formato SDI que se entregaron a los servidores de streaming son convertidas a formato Real y se almacenan en el servidor de vídeo para Internet (SENCPI1). Cuando un usuario solicita a acceder a la sesión en directo, el sistema muestra una pantalla en la que se mezclan ficheros XML tratados con hojas de estilo XSL y el vídeo propiamente dicho, mostrándose una pantalla como la que aparece en la figura 2.

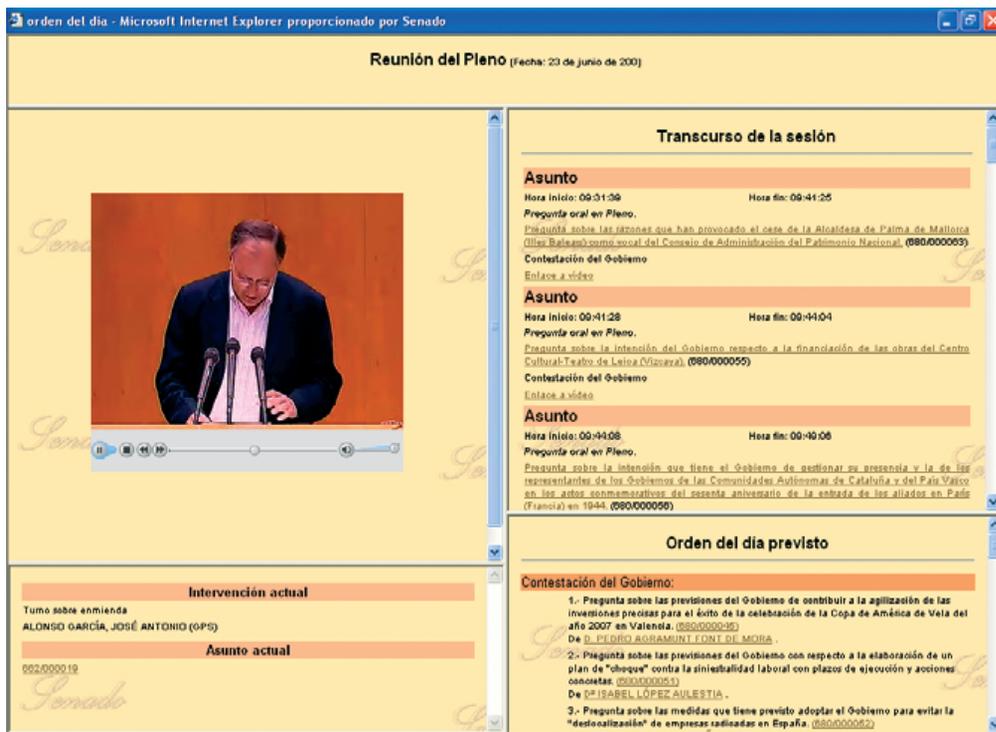


Figura 2

PROCESO DE TRANSCODIFICACIÓN Y PUBLICACIÓN

Una vez terminada la sesión que se transmite en directo, es necesario realizar un proceso de post producción. El sistema Tarsys facilita una pantalla para modificar los metadatos asociados a los cortes de vídeo en el proceso de catalogación en directo.

Tras completar los metadatos de la sesión se dispone de una serie de ficheros en formato real de 30 minutos cada uno con la sesión completa, y de un fichero con todos los metadatos

convenientemente validados. Un operador debe asociar los ficheros de vídeo con los metadatos, tras lo que se inicia el proceso de transcodificación definitiva de los materiales en formato MPEG-1 a Real. Este proceso almacena los ficheros de vídeo en la estación S.A.N. Por otro lado, se generan y traspasan al servidor de ficheros del subsistema Gelabert diferentes ficheros con extensión smi, ram y xml, que contienen datos y otra información necesaria para servir los vídeos y la información complementaria en Internet. Este traspaso usa el programa CGI del que ya se ha hablado en el presente documento. Este proceso se ilustra en el esquema del anexo III.

CONSULTA AL ARCHIVO DE IMAGEN Y SONIDO DESDE TARSYS

El sistema adquirido incluye un número limitado de licencias de la interfaz de consulta de Tarsys (ver figura 3).

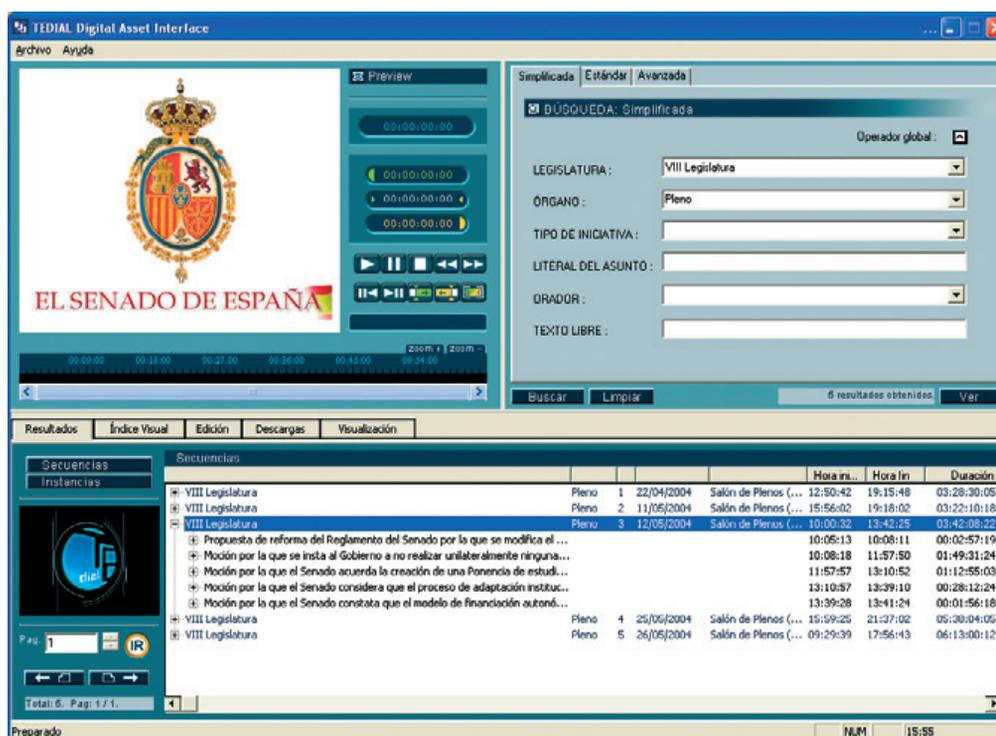


Figura 3

Mediante tres tipos de pantallas (simplificada, estándar y avanzada) se pueden formular consultas combinando distintos criterios de búsqueda para obtener los materiales deseados. Este interfaz es el adecuado para recuperar material de alta calidad para su entrega a los medios de comunicación, o para su reproducción en formato VHS, CD-R p DVD.

Consulta al archivo de imagen y sonido desde la web del Senado. El acceso a la información multimedia de las sesiones que se facilita a todos los usuarios del Senado y a todos los conectados a Internet, es el que se incluye en la web del Senado.

Hay tres puntos de acceso. En la pantalla "Sesiones Plenarias celebradas durante la VIII Legislatura" del enlace "Sesiones plenarias" de la zona "Qué hace" de la página inicial se incluye una relación de todas las sesiones de Pleno que han tenido lugar. Además de otra información textual (orden del día, diario de sesiones y resultado de la tramitación) se ofrece un enlace a la pantalla de la sesión, que permitirá acceder a cada asunto e intervención concreta (ver figura 4).

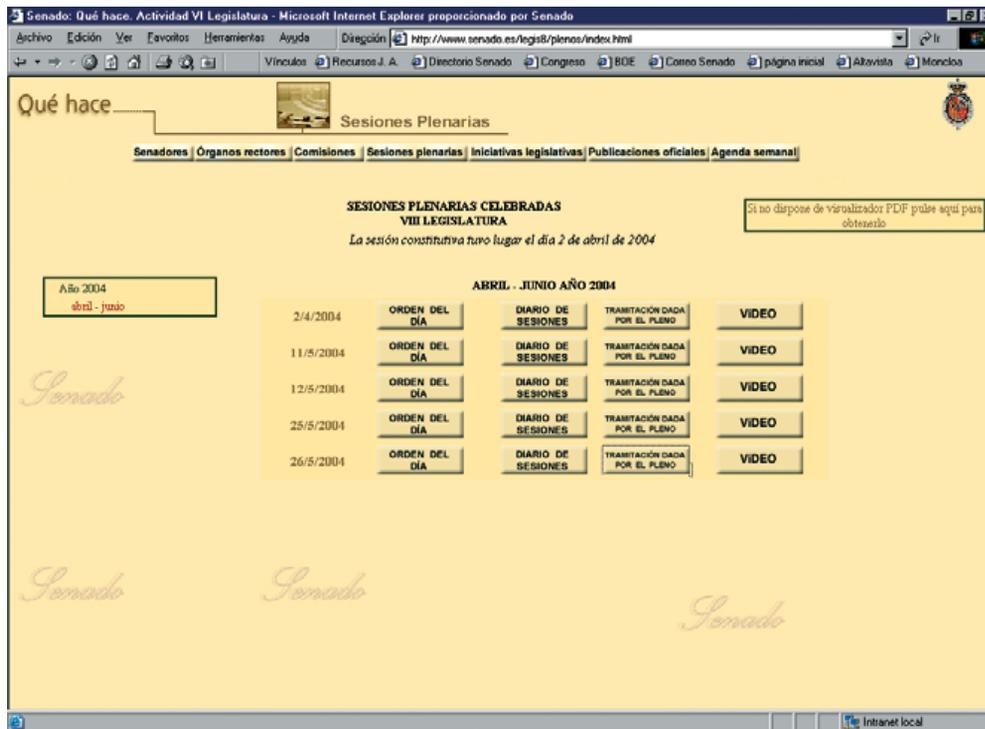


Figura 4

Desde los enlaces de iniciativas legislativas y de control se accede a las fichas de los expedientes (ver figura 5), en los que se incluye un botón que enlaza con una lista de vídeos de intervenciones concretas.

En los asuntos que no se catalogan a nivel de intervención, preguntas en Pleno, por ejemplo, el enlace en la ficha del expediente llevará a visualizar directamente el vídeo del asunto en su totalidad.

Por último, las fichas de los Senadores, a las que se accede desde la zona "Qué es" de la página inicial, incorporan también enlaces a los vídeos con sus respectivas intervenciones.

El anexo IV recoge un esquema de funcionamiento del subsistema que permite servir los "cortes" de vídeo deseados. Aparecen en el mismo el servidor de streaming, la estación de almacenamiento S.A.N., de la que aquel obtiene los vídeos, y el servidor de ficheros desde el que se obtienen los necesarios para mostrara la información complementaria en pantalla.

Requisitos de los ordenadores que accedan al archivo de imagen y sonido del Senado. Es necesario disponer de un ordenador con suficiente potencia que tenga instalado el visualizador

de la empresa Real Networks, que se puede obtener de Internet. Los detalles del proceso y otra información técnica de interés se publican en la página http://www.senado.es/canal_senado/ayudavideo.html de la web del Senado. Se aconseja disponer de, al menos, la capacidad de comunicación que permiten las líneas ADSL.



Figura 5

ANEXO I

Contenido del fichero directo_XXXXXXX_yyy_zz.xml

```

<?xml version="1.0" encoding="ISO8859-1"?>
<?xml-stylesheet type="text/xsl" href="http://www.senado.es/plantillas/directo.xsl"?>
<main>
<sesion>
<metadatos>
<fecha_inicio><![CDATA[25/05/2004 15:59:25]]></fecha_inicio>
<fecha_prevista_inicio><![CDATA[25/05/2004 16:00:00]]></fecha_prevista_inicio>
<organo><![CDATA[Pleno]]></organo>
<codigo_organo><![CDATA[S000040]]></codigo_organo>
<numero_sesion><![CDATA[4]]></numero_sesion>
<numero_convocatoria><![CDATA[1]]></numero_convocatoria>
<tipo_sesion><![CDATA[Pleno]]></tipo_sesion>
<codigo_sesion><![CDATA[0]]></codigo_sesion>
<sala><![CDATA[Salón de Plenos (nuevo)]]></sala>
<codigo_sala><![CDATA[1]]></codigo_sala>
</metadatos>
<asunto>
<metadatos>
<fecha_inicio><![CDATA[25/05/2004 20:14:04]]></fecha_inicio>
<fecha_fin><![CDATA[25/05/2004 21:36:25]]></fecha_fin>
<fecha_prevista_inicio><![CDATA[25/05/2004 16:00:00]]></fecha_prevista_inicio>
<numero_orden><![CDATA[5]]></numero_orden>
<tipo_iniciativa><![CDATA[Moción ante el Pleno.]]></tipo_iniciativa>
<codigo_iniciativa><![CDATA[662]]></codigo_iniciativa>
<numero_iniciativa><![CDATA[000009]]></numero_iniciativa>
<literal_iniciativa><![CDATA[Moción por la que se insta a la Cámara a impulsar la reforma
del Reglamento del Senado con el fin de hacer posible la utilización de cualquiera de las lenguas
que tienen carácter oficial en alguna de las Comunidades Autónomas, así como a que el Gobierno
fomente el plurilingüismo en la Administración General del Estado y ante la Unión Euro-
pea.]]></literal_iniciativa>
<evento_sesion><![CDATA[Sesión del Pleno]]></evento_sesion>
<codigo_evento><![CDATA[061]]></codigo_evento>
<resultado_tramitacion><![CDATA[Aprobado]]></resultado_tramitacion>
<numero_orden_previsto><![CDATA[5]]></numero_orden_previsto>
</metadatos>
</asunto>
</sesion>
</main>
Esquema del contenido de los ficheros diferido_temp_XXXXXXX_yyy_zz.xml y
diferido_fin_XXXXXXX_yyy_zz.xml

```

```

<sesion>
  <metadatos>
    /** Aquí los metadatos asociados a la sesión **/
  </metadatos>
<asunto>
  <smil>
    /** Aquí la entrada smil para el asunto nº 1 **/
  </smil>
  <metadatos>
    /** Aquí los metadatos asociados al asunto nº 1 **/
  </metadatos>
  <intervención> /* Intervención nº 1 del asunto nº 1 **/
    <smil>
      /** Entrada smil de la intervención nº 1 del asunto nº 1 **/
    </smil>
    <metadatos>
  /** Metadatos de la intervención nº 1 del asunto nº 1 **/
    </metadatos>
  </intervención>
<intervención> /* Intervención nº 2 del asunto nº 1 **/
  <smil>
    /** Entrada smil de la intervención nº 2 del asunto nº 1 **/
  </smil>
  <metadatos>
  /** Metadatos de la intervención nº 2 del asunto nº 1 **/
    </metadatos>
  </intervención>
  .....
</asunto> /** Fin del asunto nº 1 **/
<asunto>
  ..... información del asunto nº 2.....
</asunto> /** Fin del asunto nº 2 **/
.....
<asunto>
  ....información del asunto nº n
  (anterior al que se está produciendo en este momento).....
</asunto> /** Fin del asunto nº n **/

```

Contenido de ficheros .ram

Contiene únicamente una referencia al correspondiente fichero SMIL del corte de vídeo deseado

rtsp://helix.senado.es/realsan/ses_fin_S000040_005_01.smi

Contenido de ficheros SMIL.

Contiene comandos en formato SMIL que indican que ficheros de vídeo reproducir y, dentro de cada uno de ellos, el momento del comienzo y final de la reproducción.

```
<smil>  
    <video src="C:\\tmp\\asunto.rm" begin="03:14.0" end="06:23.0"/>  
</smil>
```

Denominación de los ficheros transmitidos de Tarsys a Gelabert

En los nombres de los ficheros se incluyen códigos y números que significan:

xxxxxxx código del órgano

yyy número de sesión

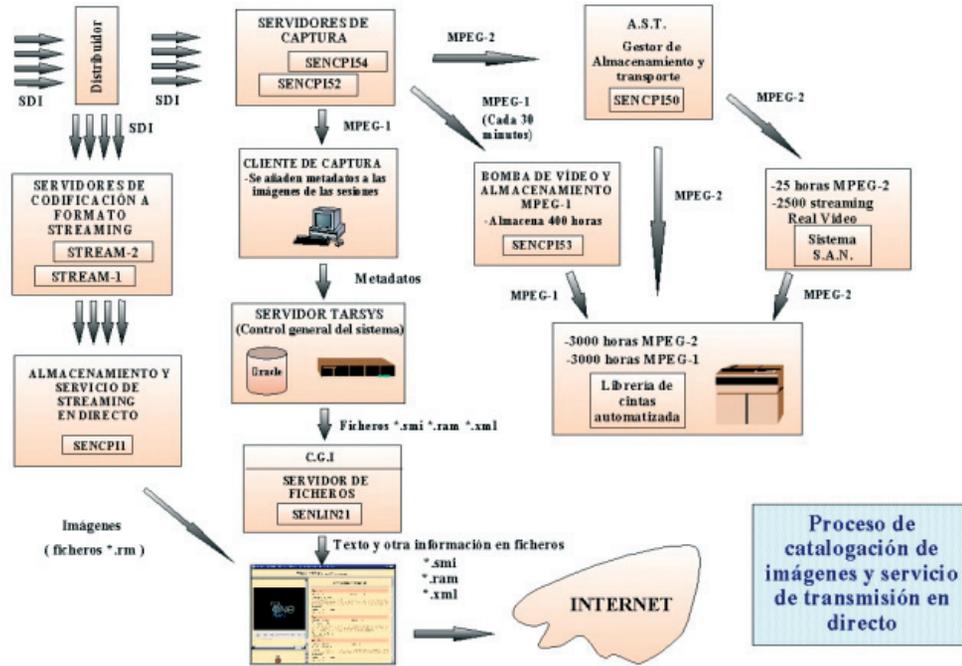
zz número de convocatoria

aaa número de orden

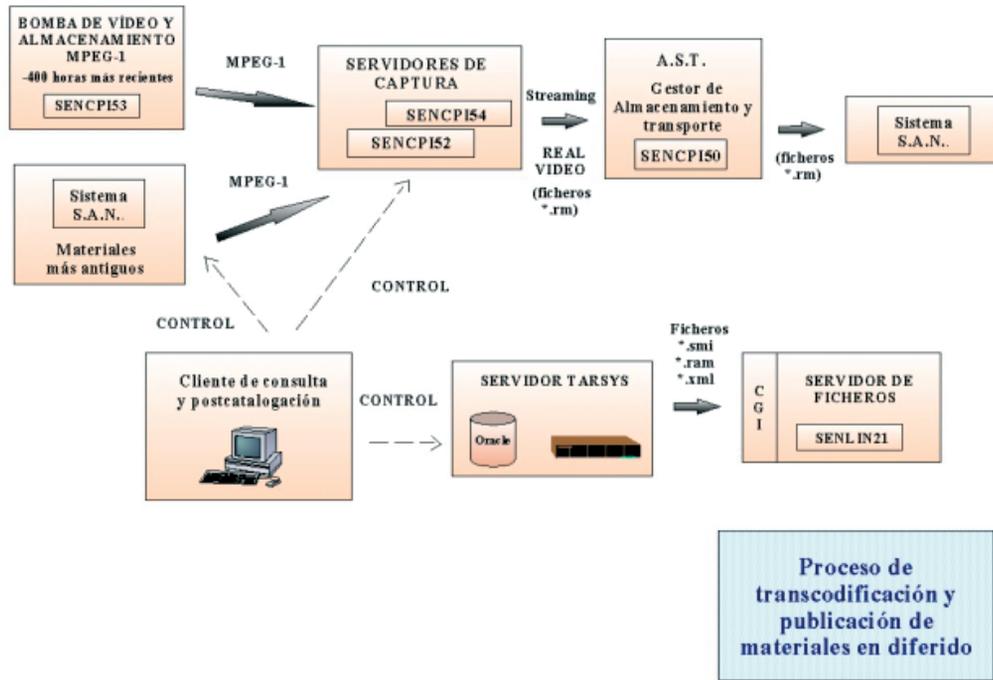
bbb número de intervención

nnn número de secuencia que identifica unívocamente al fichero rm dentro de los varios que pueden dedicarse a una sesión, asunto o intervención.

ANEXO II



ANEXO III



ANEXO IV

